HR Analytics & Employee Attrition Analysis

dimpybangoriya@gmail.com

Introduction

Employee attrition is a critical challenge for organizations, impacting workforce stability, productivity, and operational costs. Understanding the factors influencing employee turnover can help businesses implement proactive strategies to retain top talent.

In this project, I conducted Exploratory Data Analysis (EDA) in Python to analyze key factors contributing to employee attrition. Using various statistical techniques and visualizations, I examined attributes such as age, job role, salary, work-life balance, job satisfaction, and years at the company to uncover patterns and insights.

By leveraging pandas, numpy, seaborn, and matplotlib, I identified trends and correlations that can help HR teams improve employee engagement, reduce attrition rates, and make data-driven decisions for workforce management.

Missing Value & Statistics Analysis

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
hr=pd.read_csv('HR-Employee-Attrition.csv')
```

```
hr.head()
```

...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	Ye
...	1	80	0	8	0	1	6	4	
...	4	80	1	10	3	3	10	7	
...	2	80	0	7	3	3	0	0	
...	3	80	0	8	3	3	8	7	
...	4	80	1	6	3	3	2	2	

```
hr.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
```

Null Value Analysis

[56]:
```
hr.isnull()
```

[56]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfa
0	False	False	False	False	False	False	False	False	False	False	...	
1	False	False	False	False	False	False	False	False	False	False	...	
2	False	False	False	False	False	False	False	False	False	False	...	
3	False	False	False	False	False	False	False	False	False	False	...	
4	False	False	False	False	False	False	False	False	False	False	...	
...	...	...	...	...	...	...	...	...	...	...	...	
1465	False	False	False	False	False	False	False	False	False	False	...	
1466	False	False	False	False	False	False	False	False	False	False	...	
1467	False	False	False	False	False	False	False	False	False	False	...	
1468	False	False	False	False	False	False	False	False	False	False	...	
1469	False	False	False	False	False	False	False	False	False	False	...	

1470 rows × 35 columns

[ ]:

[58]:
```
hr.isnull().sum()
```

Null Values Count

[58]:
```
hr.isnull().sum()
```

[58]:
```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
```

Summary Statistics

```
[ ]:
```

```
[60]: hr.describe()
```

[60]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLev
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470.000000	1470.000000	1470.00000
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306	2.721769	65.891156	2.729932	2.06394
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335	1.093082	20.329428	0.711561	1.10694
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	1.000000	30.000000	1.000000	1.00000
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000	2.000000	48.000000	2.000000	1.00000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000	3.000000	66.000000	3.000000	2.00000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000	4.000000	83.750000	3.000000	3.00000
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000	4.000000	100.000000	4.000000	5.00000

8 rows × 26 columns

```
[ ]:
```

```
[62]: hr.isnull().sum().sum()
```

[62]:
```
0
```

```
[ ]:
```

```
[ ]:
```

Attrition Rate Analysis

```
[ ]:
```

```
[ ]: # Attrition Rate Analysis
```

```
[ ]:
```

```
[ ]: #total attrition
```

```
[ ]:
```

```
[84]: hr['Attrition'].value_counts()
```

```
[84]: Attrition
      No     1233
      Yes     237
      Name: count, dtype: int64
```

```
[ ]:
```

```
[ ]: # Percentage of Attrition
```

```
[ ]:
```

```
[192]: attrition_rate=hr['Attrition'].value_counts(normalize=True) * 100
```

```
[194]: attrition_rate
```

```
[194]: Attrition
       No     83.877551
       Yes    16.122449
       Name: proportion, dtype: float64
```

Attrition Rate Analysis & Identifying Numerical & Categorical Column

```
# Percentage Employee Attrited in each Department
```

```
attrition_department=hr.groupby('Department')['Attrition'].value_counts(normalize=True).loc[:, 'Yes']
```

```
attrition_department
```

```
Department
Human Resources           0.190476
Research & Development    0.138398
Sales                     0.206278
Name: proportion, dtype: float64
```

```
# identifying numerical and categorical column
```

```
cat=[]
num=[]
for column in hr.columns:
    if hr[column].nunique() > 10:
        num.append(column)
    else:
        cat.append(column)
```

```
cat
```

```
['Attrition',
 'BusinessTravel',
 'Department',
 'Education',
 'EducationField',
 'EmployeeCount',
 'EnvironmentSatisfaction',
 'Gender',
 'JobInvolvement',
 'JobLevel',
```

Numerical Columns

```
 'RelationshipSatisfaction',
 'StandardHours',
 'StockOptionLevel',
 'TrainingTimesLastYear',
 'WorkLifeBalance']
```

[146]:
```
num
```

[146]:
```
['Age',
 'DailyRate',
 'DistanceFromHome',
 'EmployeeNumber',
 'HourlyRate',
 'MonthlyIncome',
 'MonthlyRate',
 'PercentSalaryHike',
 'TotalWorkingYears',
 'YearsAtCompany',
 'YearsInCurrentRole',
 'YearsSinceLastPromotion',
 'YearsWithCurrManager']
```

[ ]:

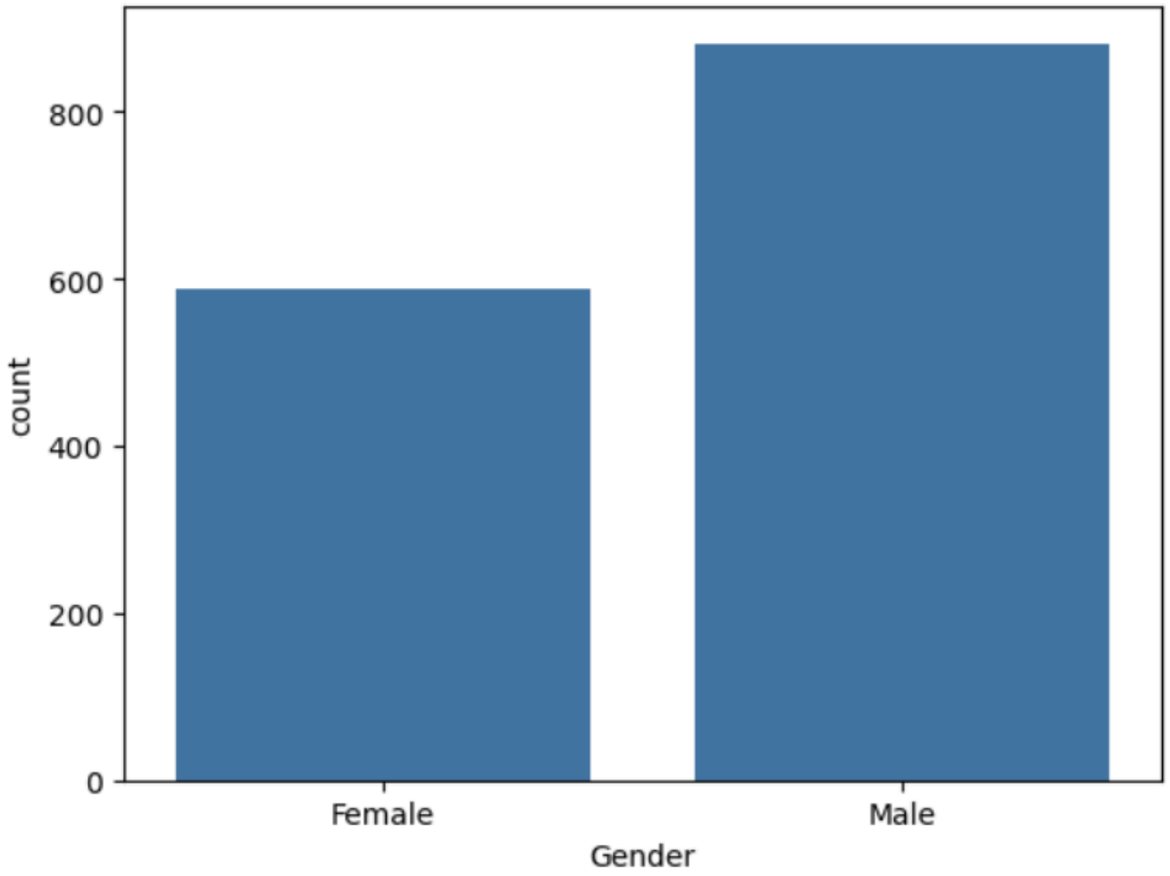
Gender Analysis Of Employee

```
[ ]:
```

```
[ ]: # Total Male & Female
```

```
[156]: sns.countplot(data=hr,x='Gender')
```

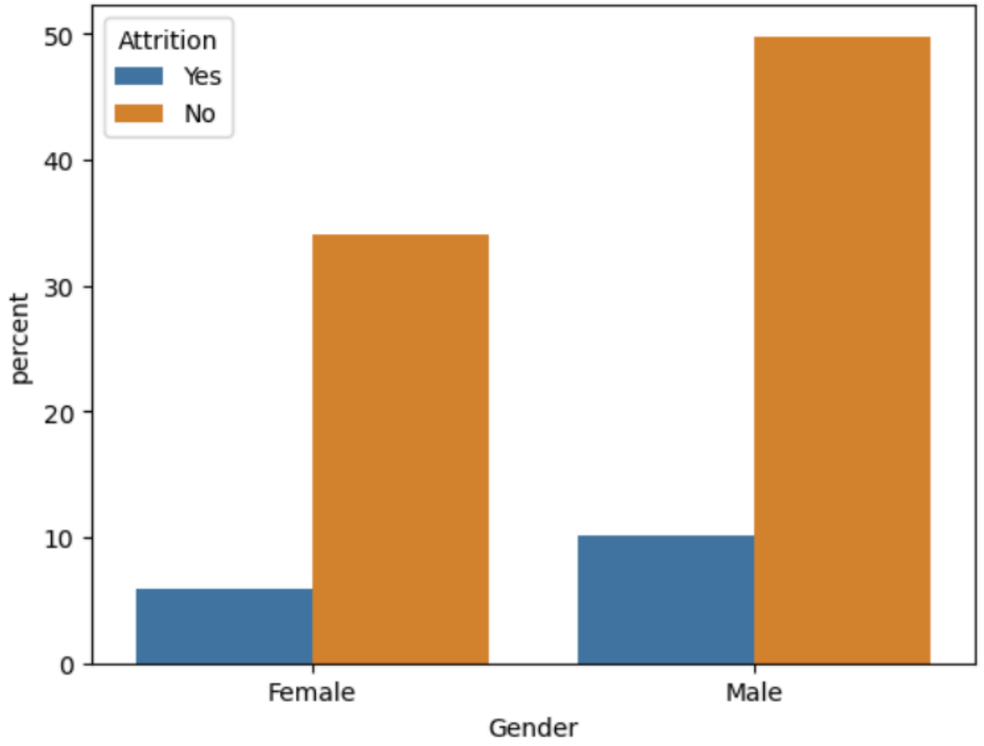
```
[156]: <Axes: xlabel='Gender', ylabel='count'>
```



Attrition Vs Gender

```
0]: sns.countplot(data=hr,x='Gender',hue="Attrition",stat='percent')
    sns.annotate(f'{p.get_height()}',
                 (p.get_x() + p.get_width() / 2., p.get_height()),
                 ha='center', va='bottom', fontsize=12, fontweight='bold', color='black')
```

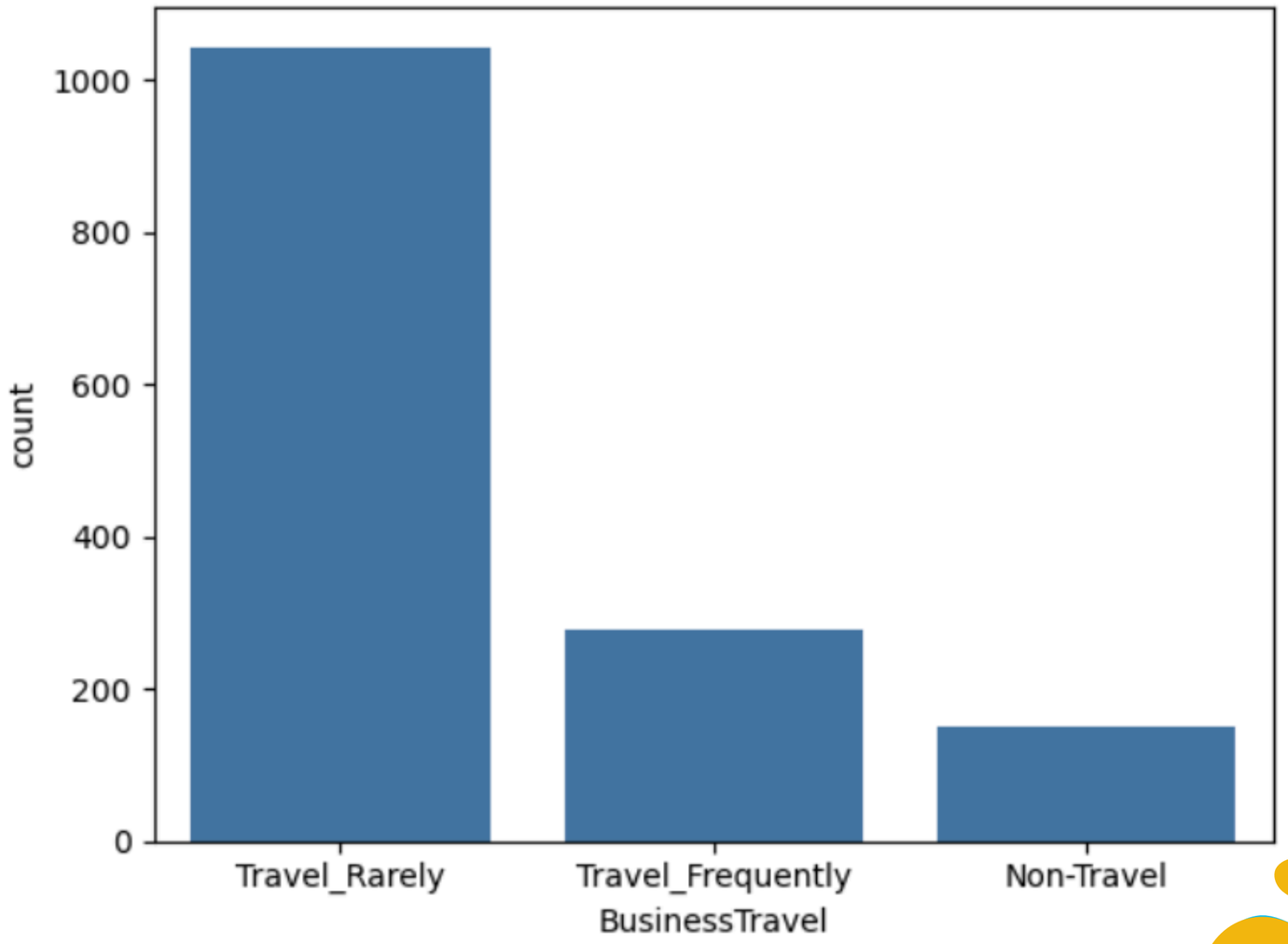
```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[200], line 2
      1 sns.countplot(data=hr,x='Gender',hue="Attrition",stat='percent')
----> 2 sns.annotate(f'{p.get_height()}',
      3                  (p.get_x() + p.get_width() / 2., p.get_height()),
      4                  ha='center', va='bottom', fontsize=12, fontweight='bold', color='black')

AttributeError: module 'seaborn' has no attribute 'annotate'
```



Statistics Of Business Travel

```
[204]: sns.countplot(data=hr,x="BusinessTravel")
```

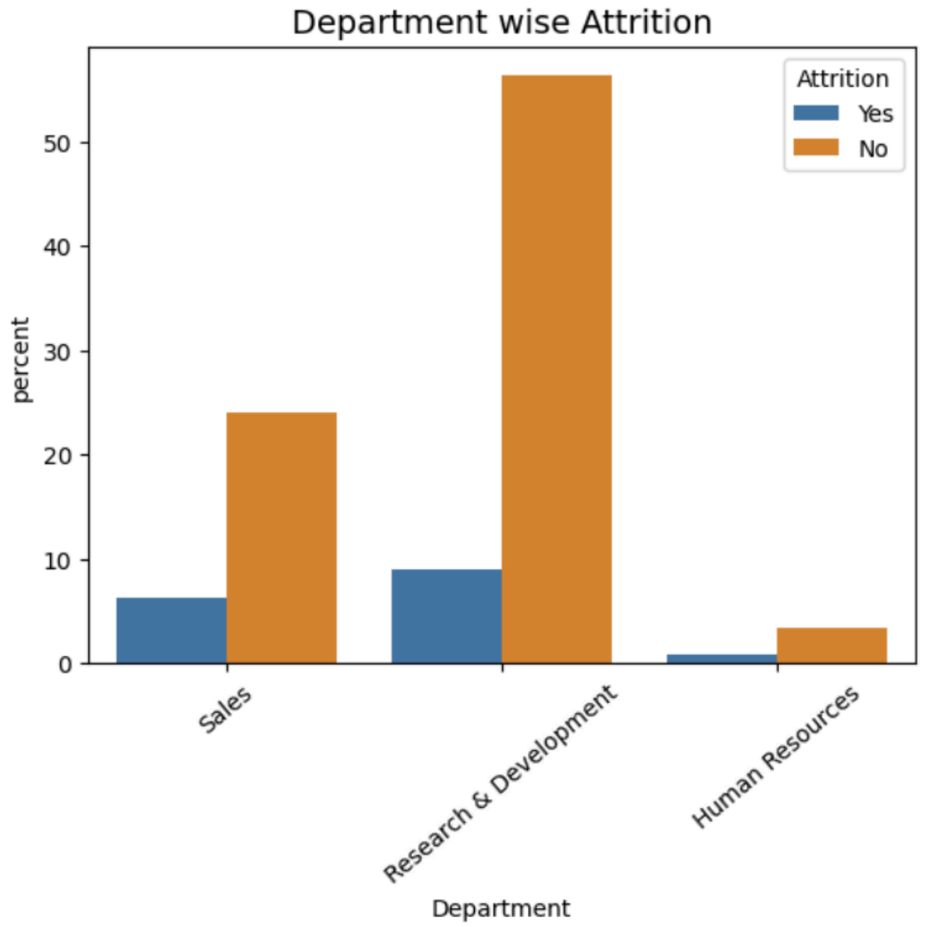
```
[204]: <Axes: xlabel='BusinessTravel', ylabel='count'>
```



Percent Attrition in Each Department

[ ]:
```
# Department Vs Attrition
```

[434]:
```
sns.countplot(data=hr,x="Department", hue="Attrition",stat='percent')
plt.xticks(rotation=40)
plt.title("Department wise Attrition", fontsize=14)
plt.show()
```



Attrition Percent By Department


```
[448]: hr.groupby('Department')['Attrition'].value_counts(normalize=True).loc[:, 'Yes']
```

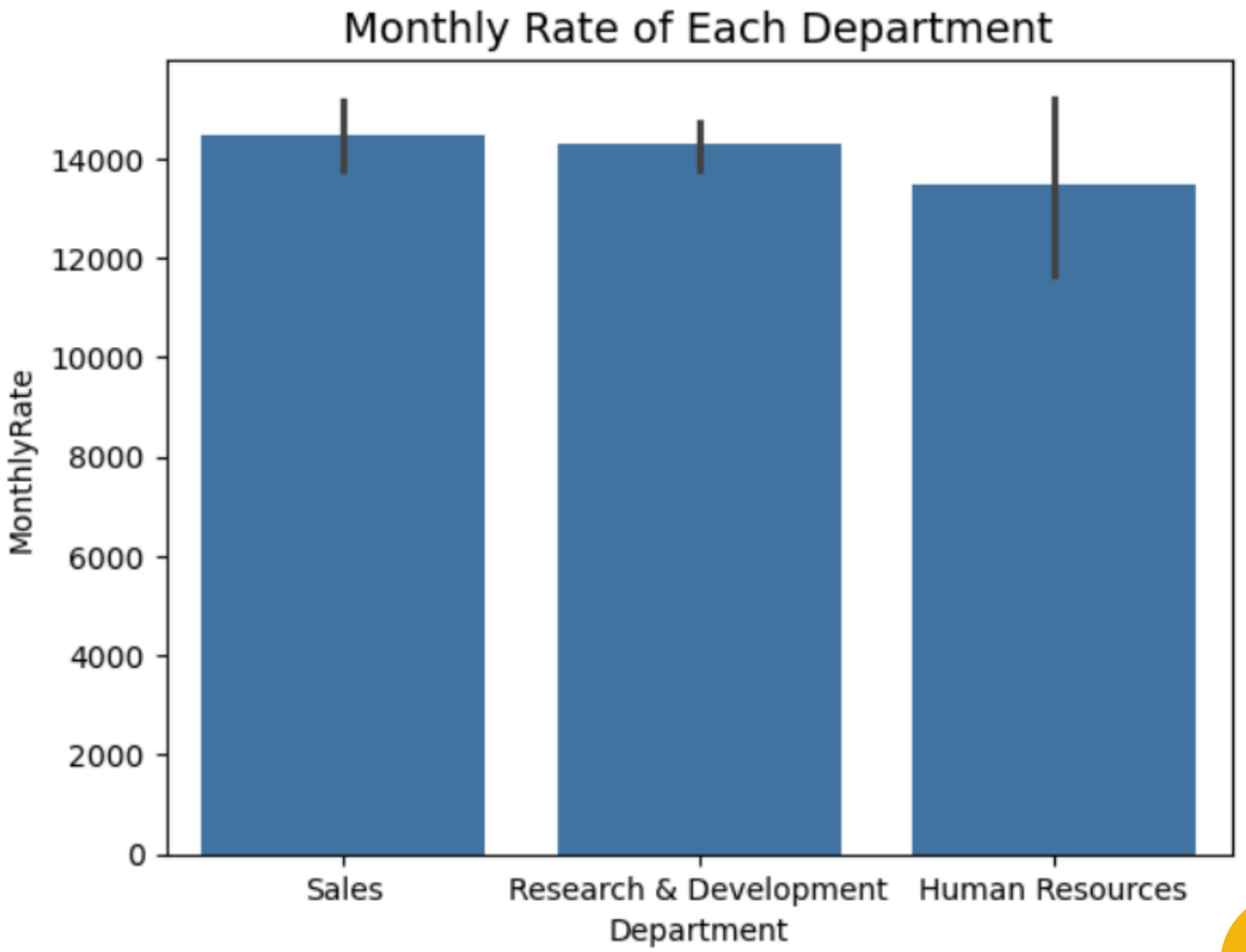
```
[448]: Department
       Human Resources           0.190476
       Research & Development    0.138398
       Sales                     0.206278
       Name: proportion, dtype: float64
```

```
[ ]:
```

Average Monthly Rate Of Employees In Each Department

[430]:
```
sns.barplot(data=hr,x='Department',y='MonthlyRate')
plt.title("Monthly Rate of Each Department", fontsize=14)
```

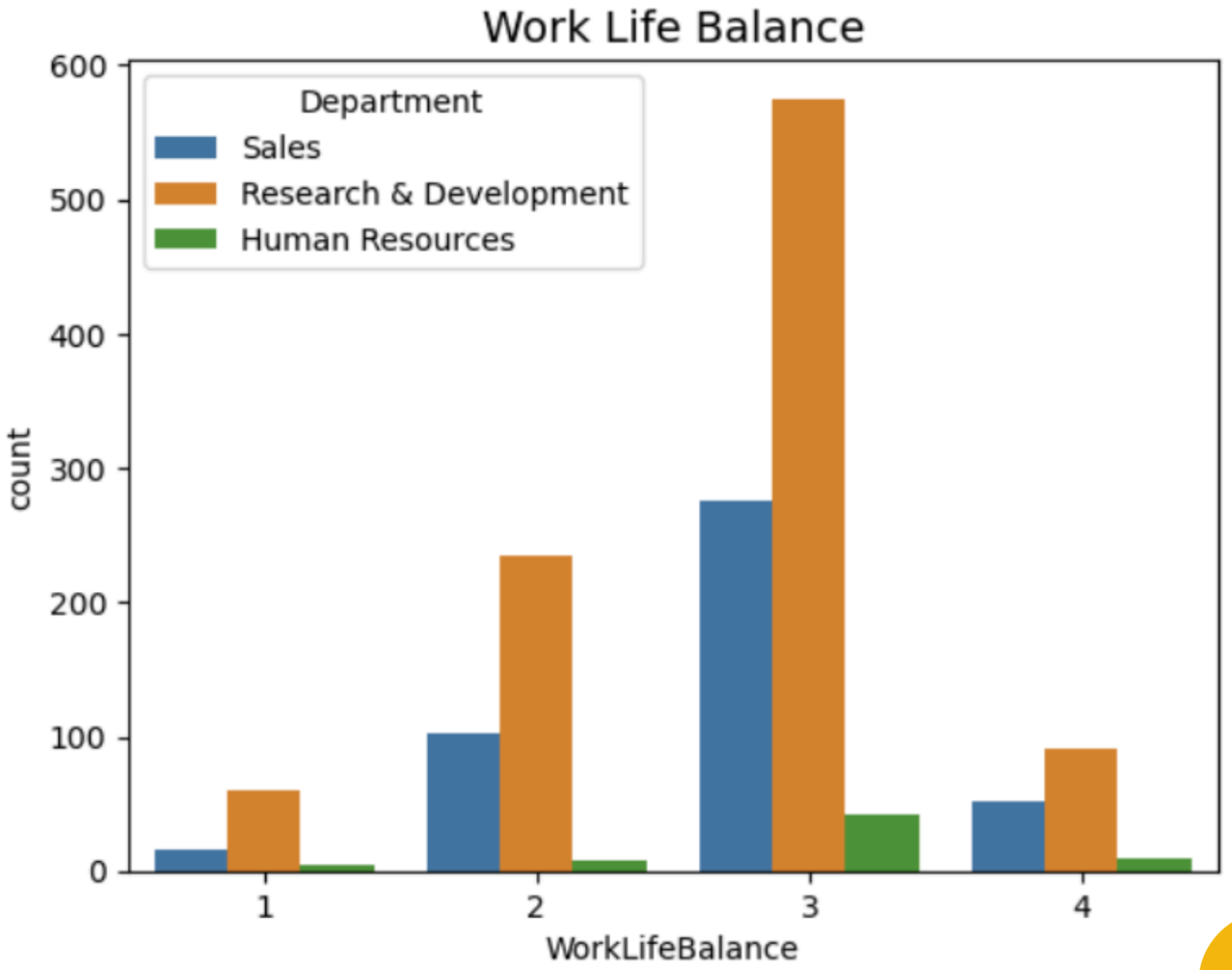
[430]:
```
Text(0.5, 1.0, 'Monthly Rate of Each Department')
```



Work Life Balance Of Employees in Each Department

```
[364]: sns.countplot(data=hr,x='WorkLifeBalance',hue='Department')
       plt.title("Work Life Balance", fontsize=14)
```

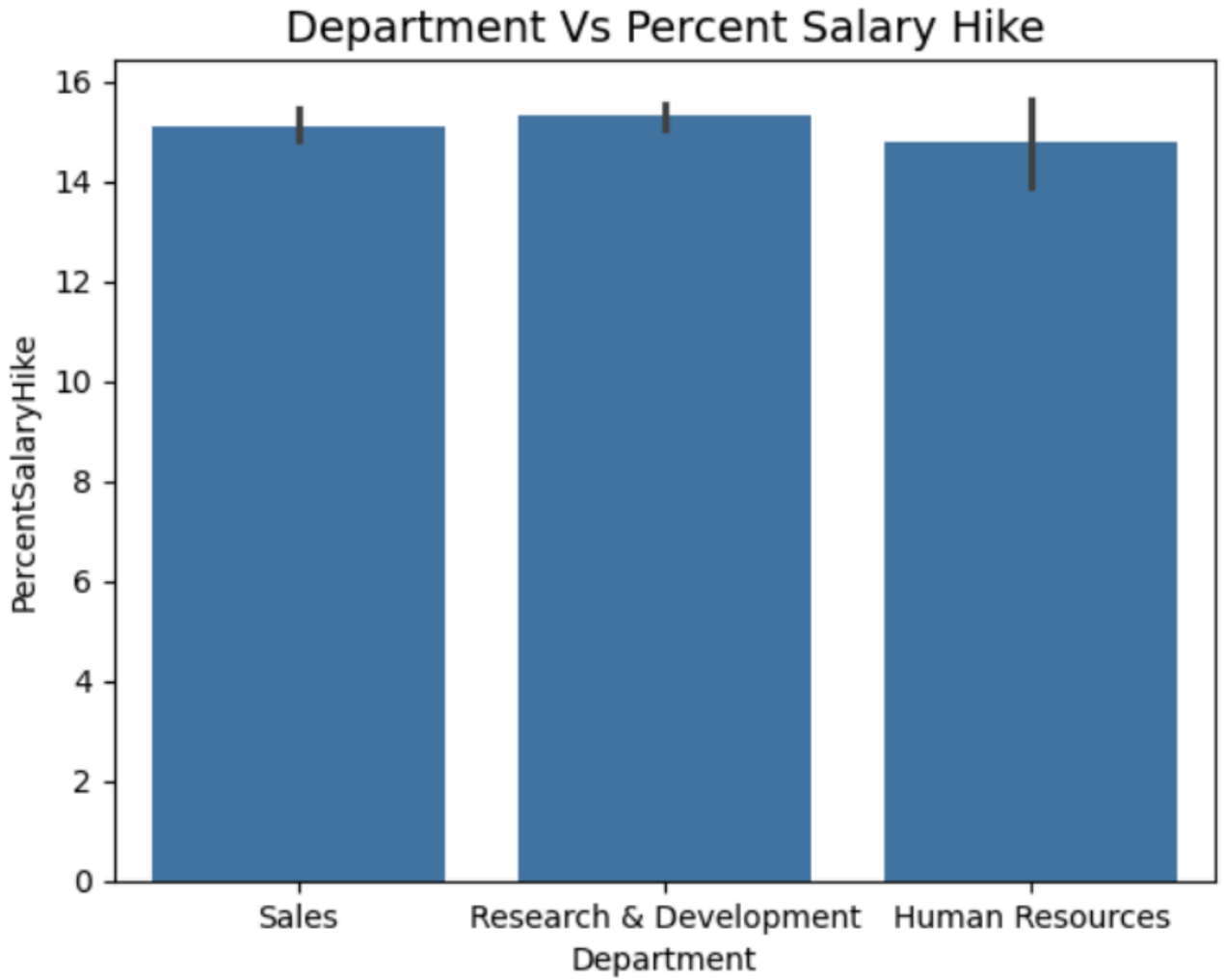
```
[364]: Text(0.5, 1.0, 'Work Life Balance')
```



Average Salary Hike in Each Department

```
[62]: sns.barplot(data=hr,x='Department',y='PercentSalaryHike',estimator='mean')
      plt.title("Department Vs Percent Salary Hike", fontsize=14)
```

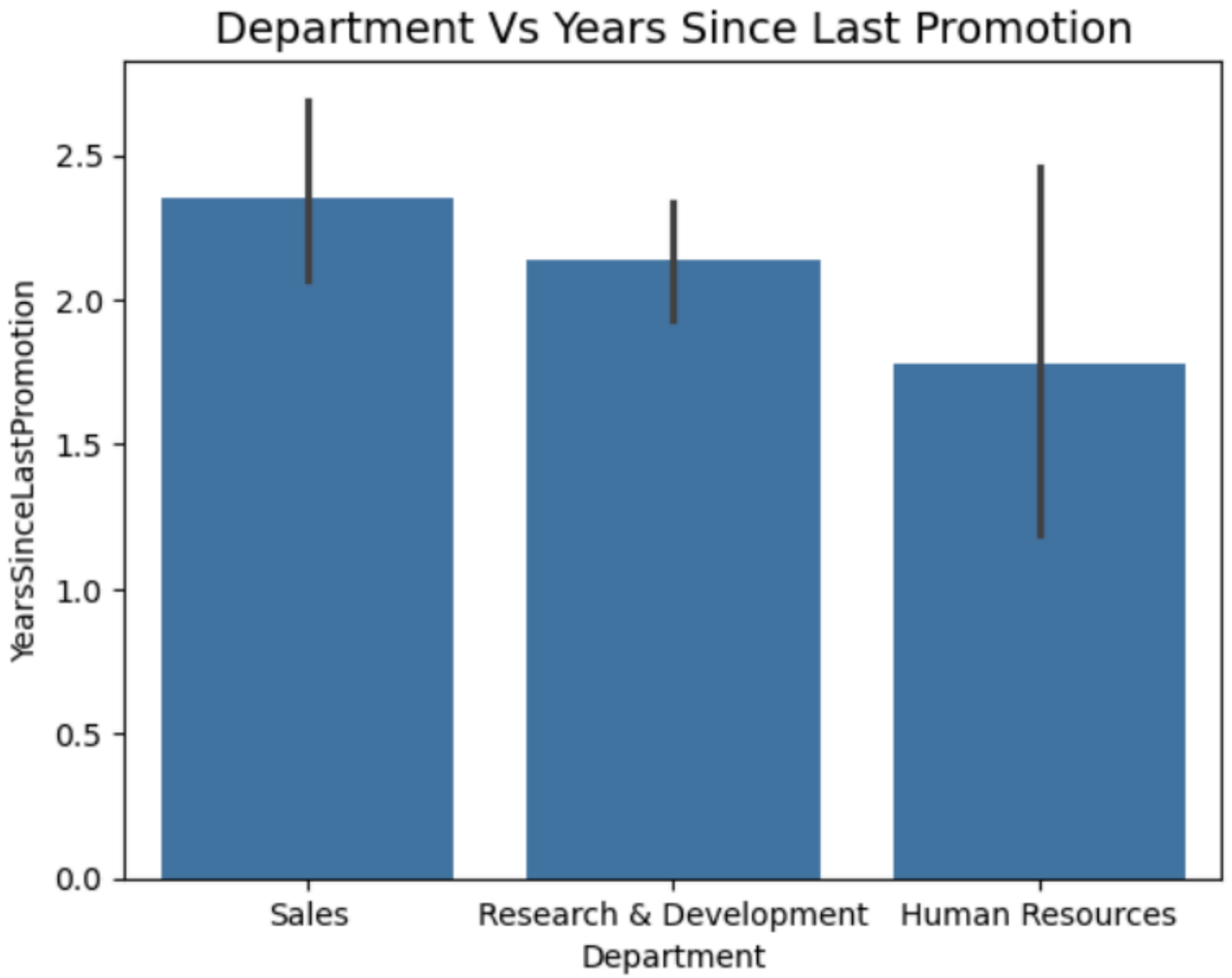
```
[62]: Text(0.5, 1.0, 'Department Vs Percent Salary Hike')
```



Average Years Since Last Promotion In Each Department

```
[356]: sns.barplot(data=hr,x='Department',y='YearsSinceLastPromotion')
       plt.title("Department Vs Years Since Last Promotion", fontsize=14)
```

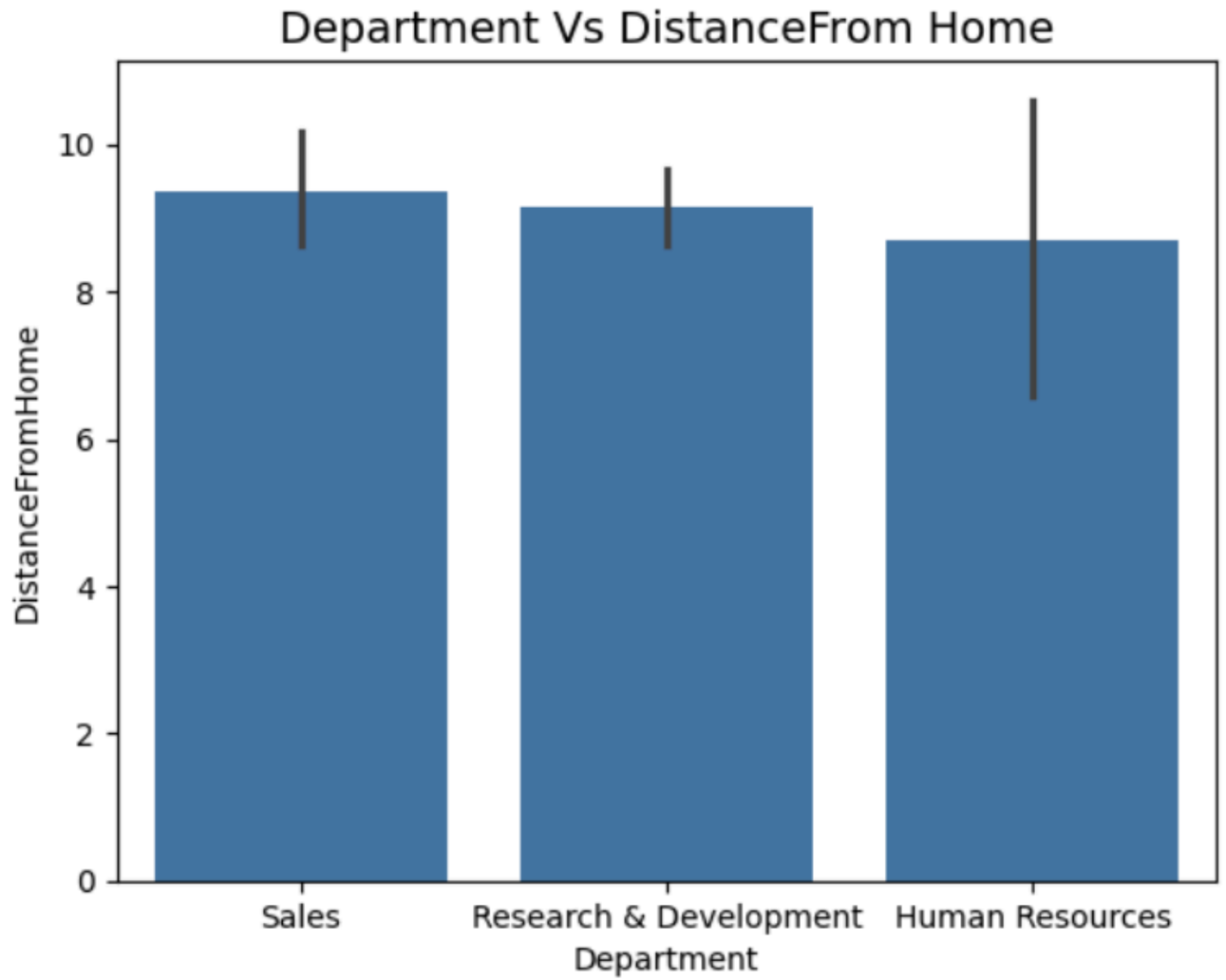
```
[356]: Text(0.5, 1.0, 'Department Vs Years Since Last Promotion')
```



Average Distance From Home Of Employees In Each Department

```
[354]: sns.barplot(data=hr,x='Department',y='DistanceFromHome')
       plt.title("Department Vs DistanceFrom Home", fontsize=14)
```

```
[354]: Text(0.5, 1.0, 'Department Vs DistanceFrom Home')
```



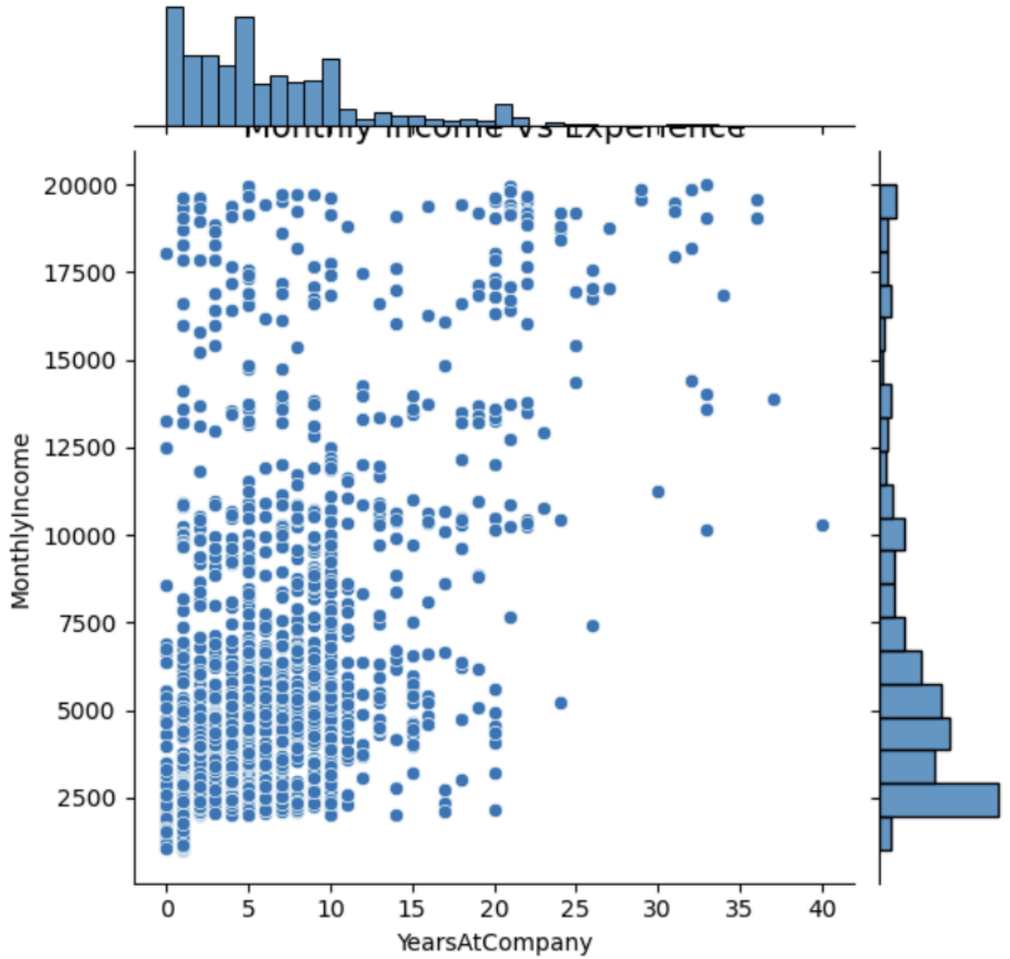
Co-Relation Between Salary & Years At Company

```
[ ]: # Years at company vs Salary
```

```
[ ]:
```

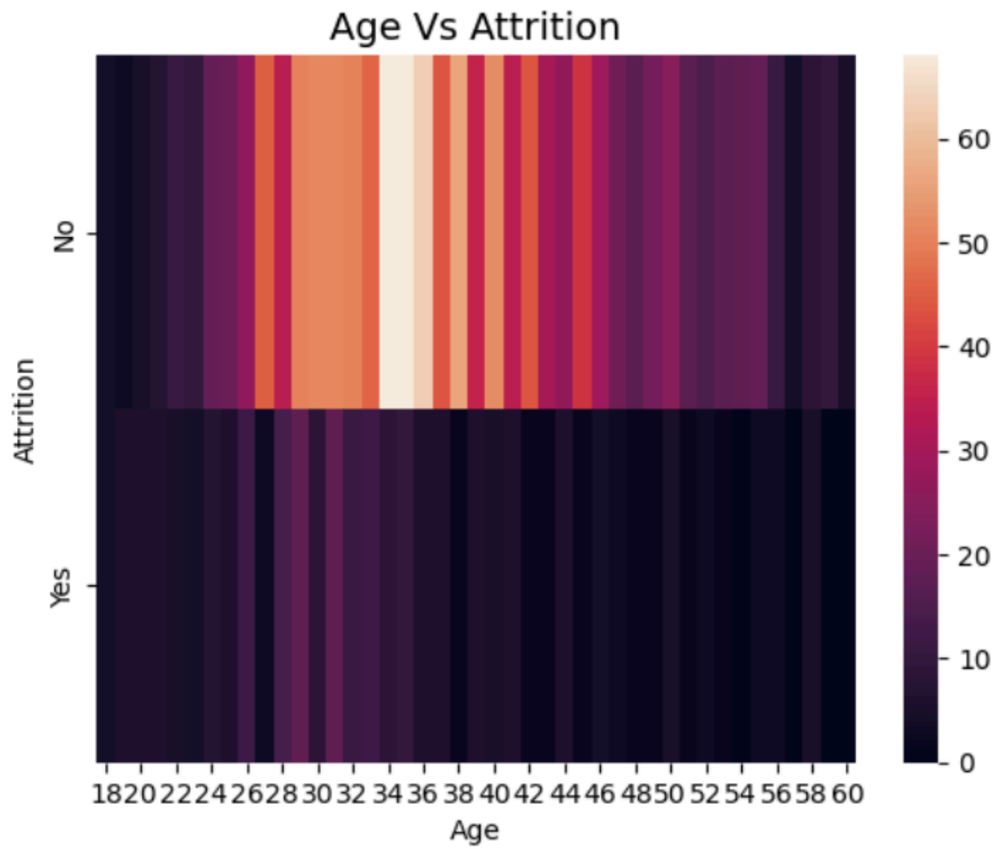
```
50]: sns.jointplot(data=hr,x='YearsAtCompany',y='MonthlyIncome')
```

```
50]: Text(0.5, 1.0, 'Monthly Income Vs Experience')
```



Co-relation Between Employee Age & Attrition

```
[ ]: # Age Vs Attrition

[ ]:

[376]: sns.heatmap(pd.crosstab(hr['Attrition'],hr['Age']))
       plt.title("Age Vs Attrition", fontsize=14)

[376]: Text(0.5, 1.0, 'Age Vs Attrition')
```




```
[ ]:
```

Work Life Balance Vs Attrition

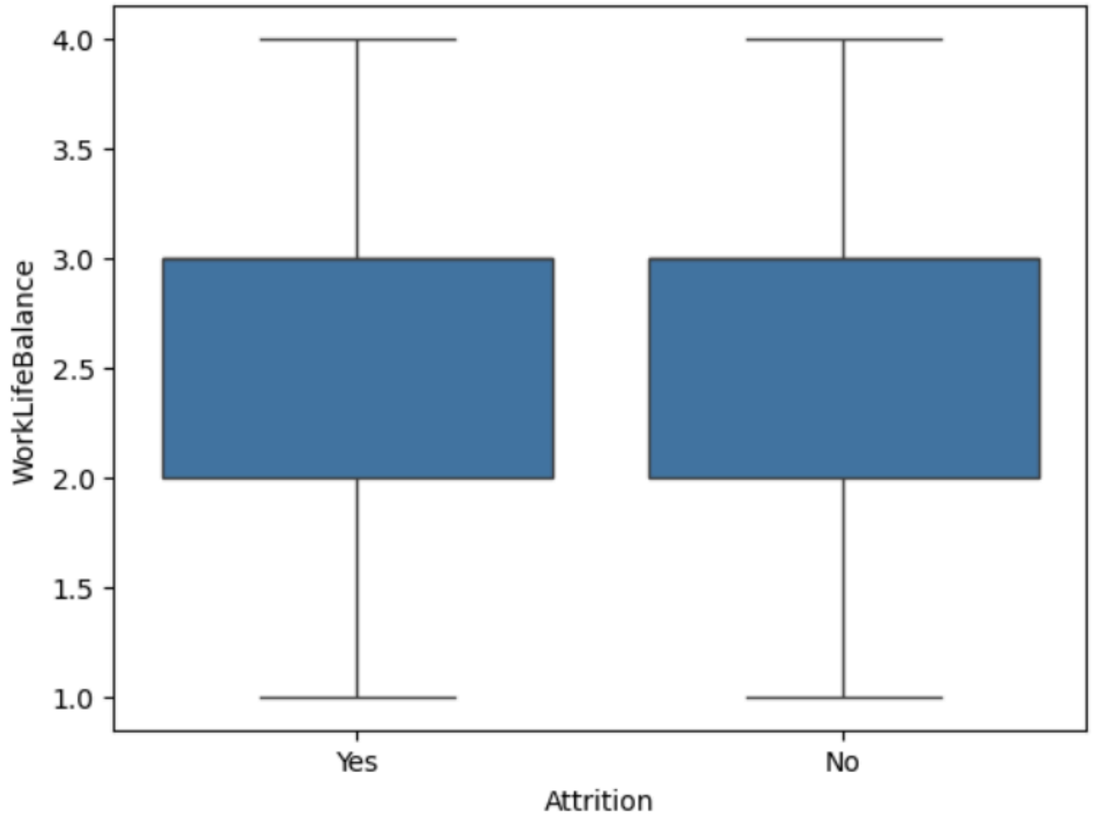
[ ]:

[ ]:
```
# Work Life balance Vs Attrition
```

[ ]:

[380]:
```
sns.boxplot(data=hr,x='Attrition',y='WorkLifeBalance')
```

[380]:
```
<Axes: xlabel='Attrition', ylabel='WorkLifeBalance'>
```



[ ]:

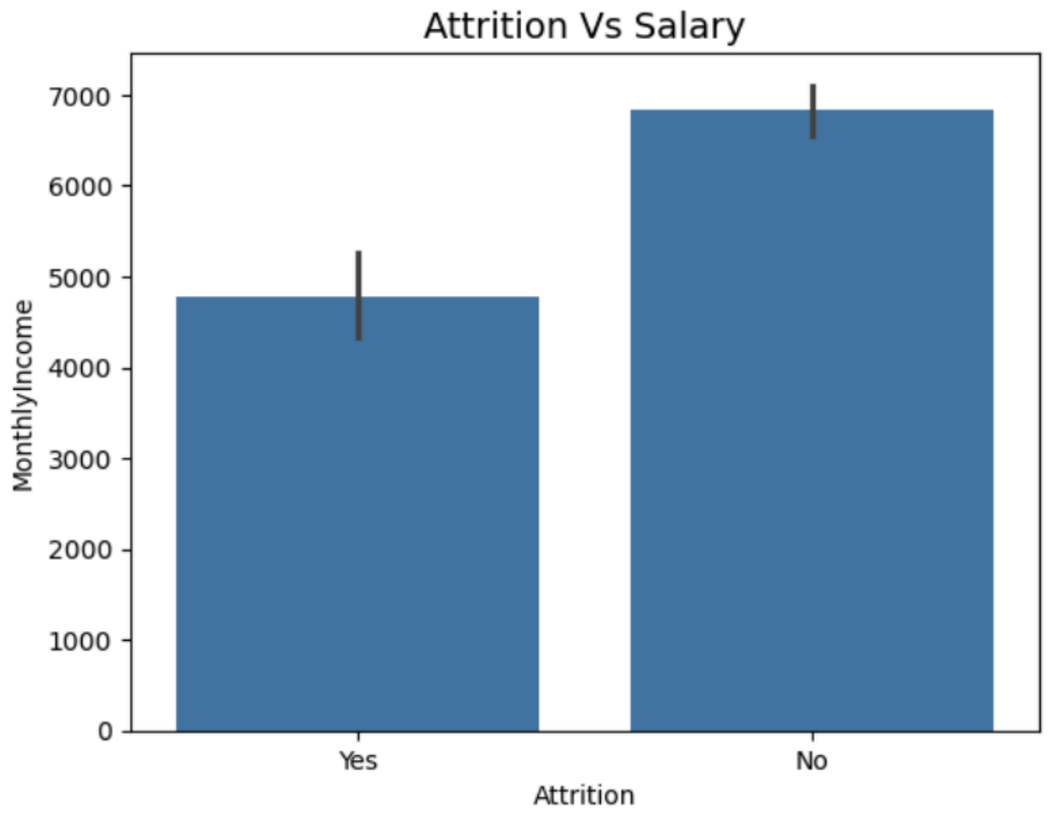
Average Monthly Income Vs Attrition

```
[ ]: # Attrition Vs Salary
```

```
[ ]:
```

```
[384]: sns.barplot(data=hr,x='Attrition',y='MonthlyIncome')
       plt.title("Attrition Vs Salary", fontsize=14)
```

```
[384]: Text(0.5, 1.0, 'Attrition Vs Salary')
```



Attrition Vs Years At Company

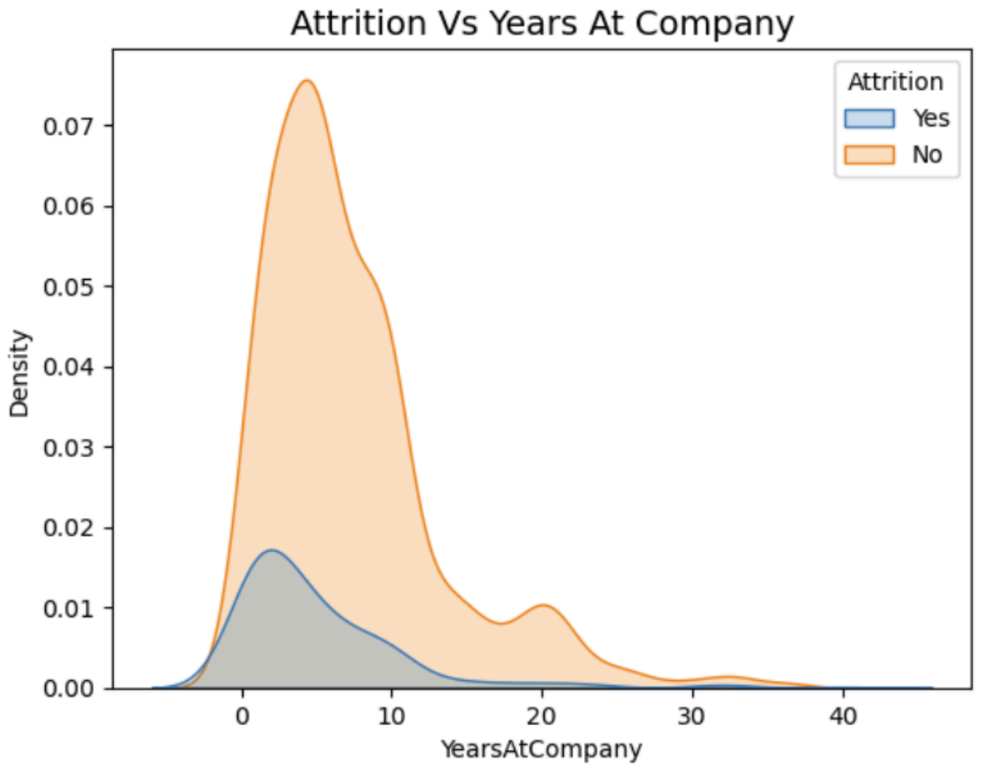
[ ]:
```
# Attrition Vs Years At Company
```

[ ]:
```

```

[436]:
```
sns.kdeplot(data=hr,x='YearsAtCompany',hue='Attrition',shade=True)
plt.title("Attrition Vs Years At Company", fontsize=14)
```

```
C:\Users\kruna\AppData\Local\Temp\ipykernel_13412\2151122389.py:1: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(data=hr,x='YearsAtCompany',hue='Attrition',shade=True)
```

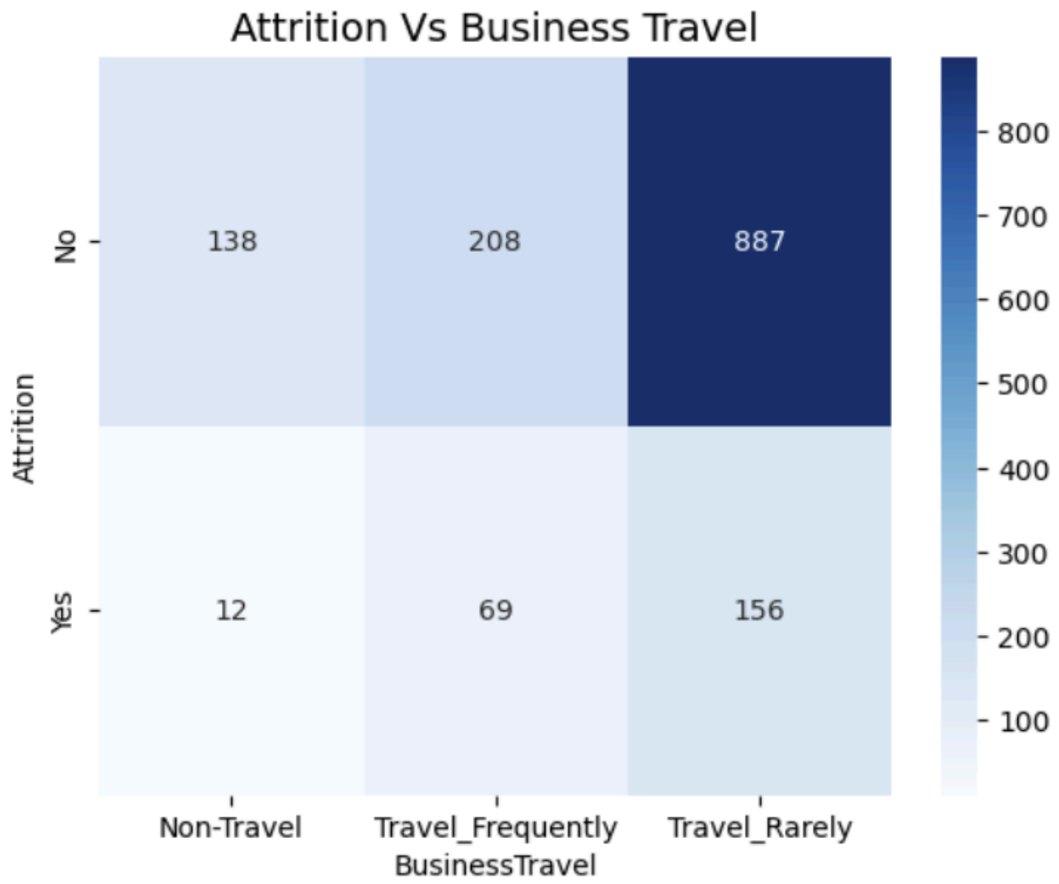
[436]:
```
Text(0.5, 1.0, 'Attrition Vs Years At Company')
```



Impact Of Business Travel On Attrition

[ ]:
```
# Attrition Vs Business Travel
```

[438]:
```
sns.heatmap(pd.crosstab(hr['Attrition'], hr['BusinessTravel']), annot=True, cmap='Blues', fmt='d')
plt.title("Attrition Vs Business Travel", fontsize=14)
plt.show()
```



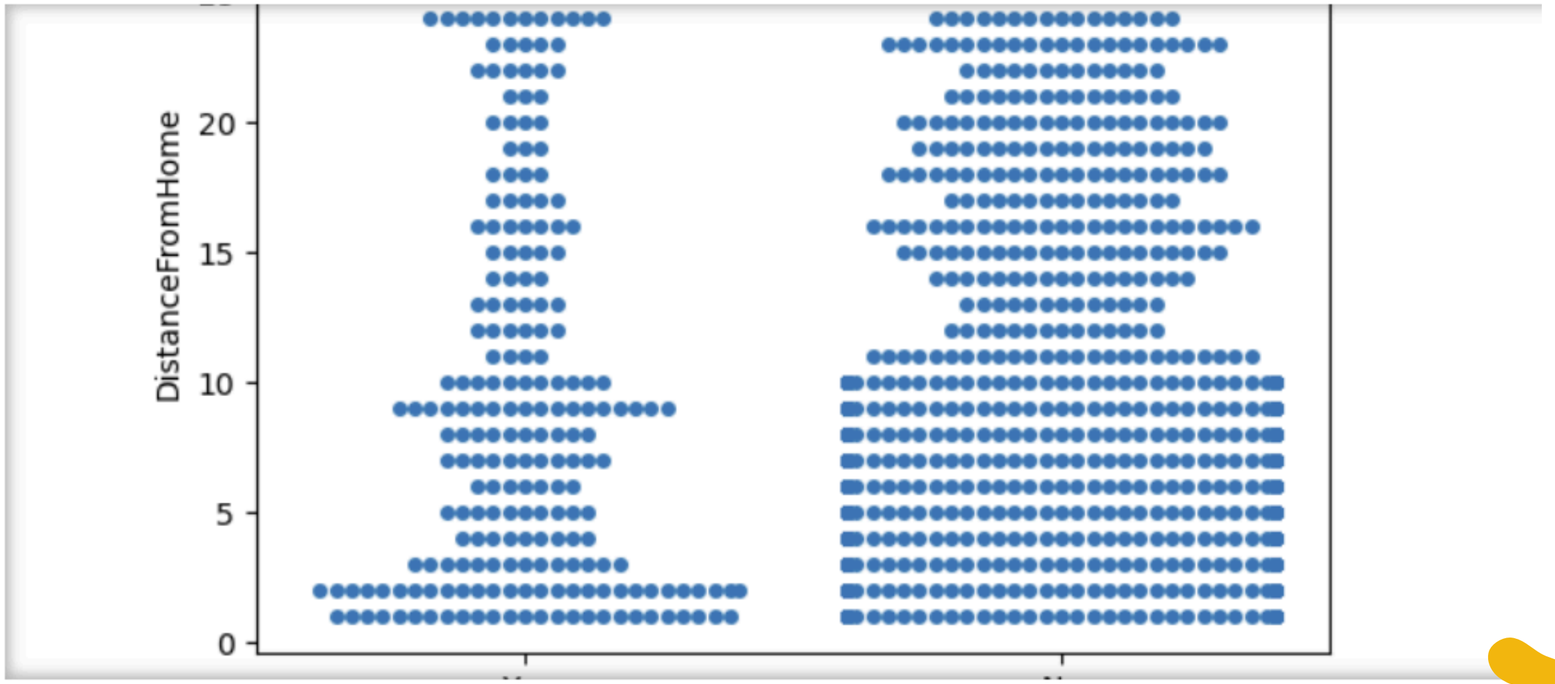
Impact of Distance From Company On Attrition

```
[ ]:
```

```
[ ]: # Attrition Vs Distance From Home
```

```
[ ]:
```

```
[42]: sns.swarmplot(x='Attrition', y='DistanceFromHome', data=hr)
plt.title("Attrition Vs Distance From Home", fontsize=14)
plt.show()
```




```
[ ]:
```

Relation Between Experience & Attrition

```
[ ]: # Attrition Vs TotalWorkingYears
```

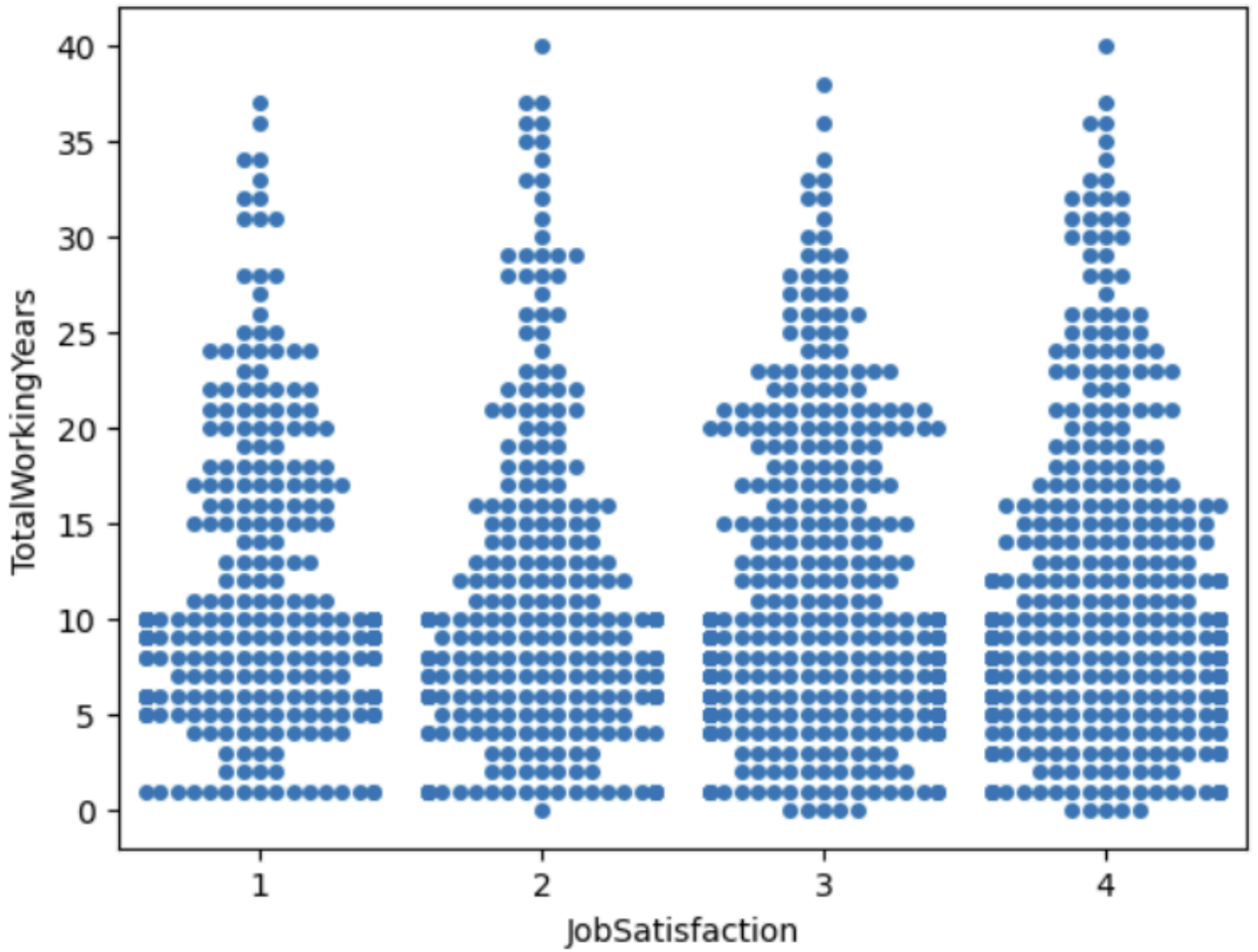
```
[ ]:
```

```
[33]: sns.violinplot(x='Attrition', y='TotalWorkingYears', data=hr, palette='pastel')
plt.title("Attrition Vs Total Working Years", fontsize=14)
plt.show()
```

```
C:\Users\kruna\AppData\Local\Temp\ipykernel_36044\3082392604.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. A
ame effect.

  sns.violinplot(x='Attrition', y='TotalWorkingYears', data=hr, palette='pastel')
```

Impact Of Job Satisfaction On Total Working Years

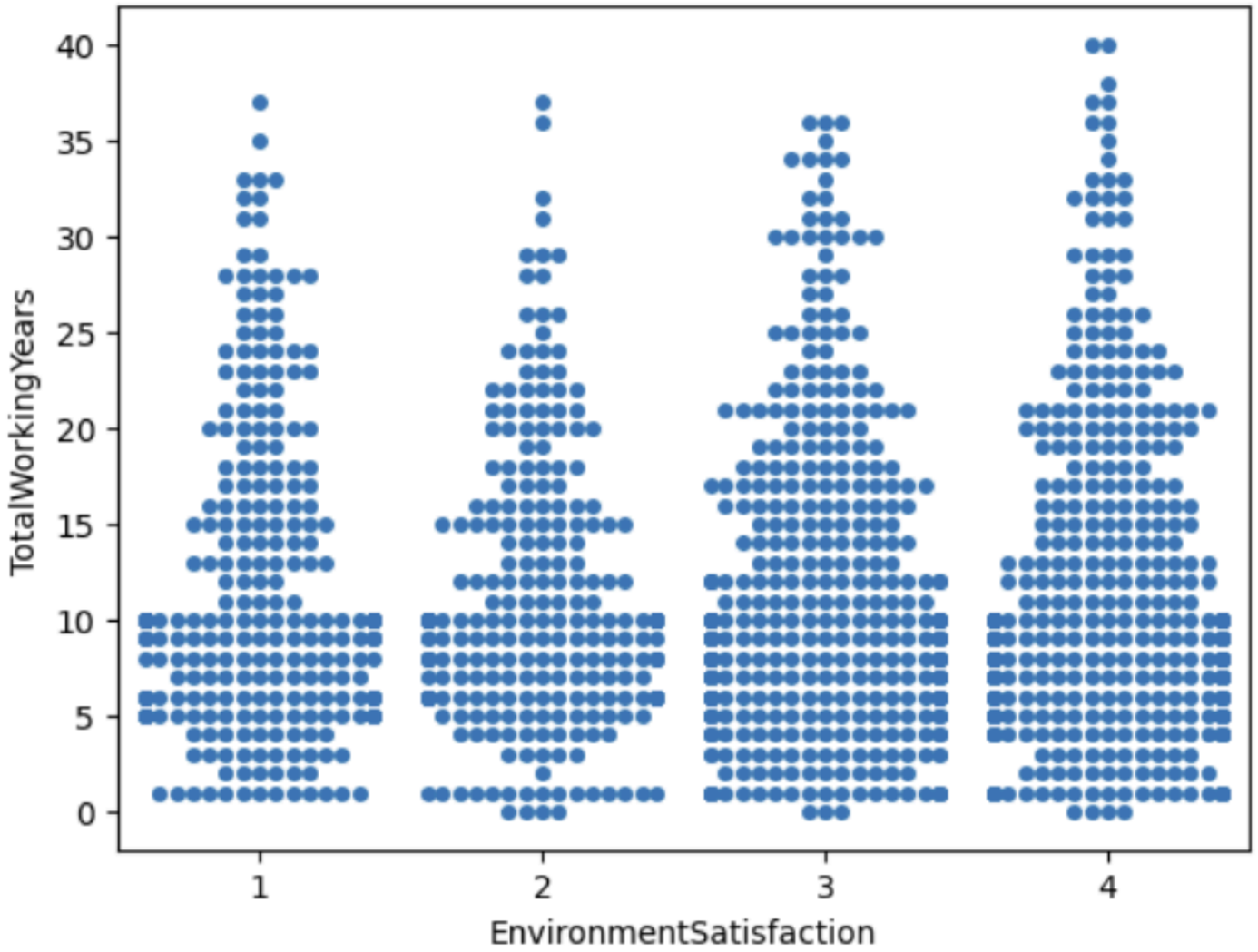
```
warnings.warn(msg, UserWarning)
```



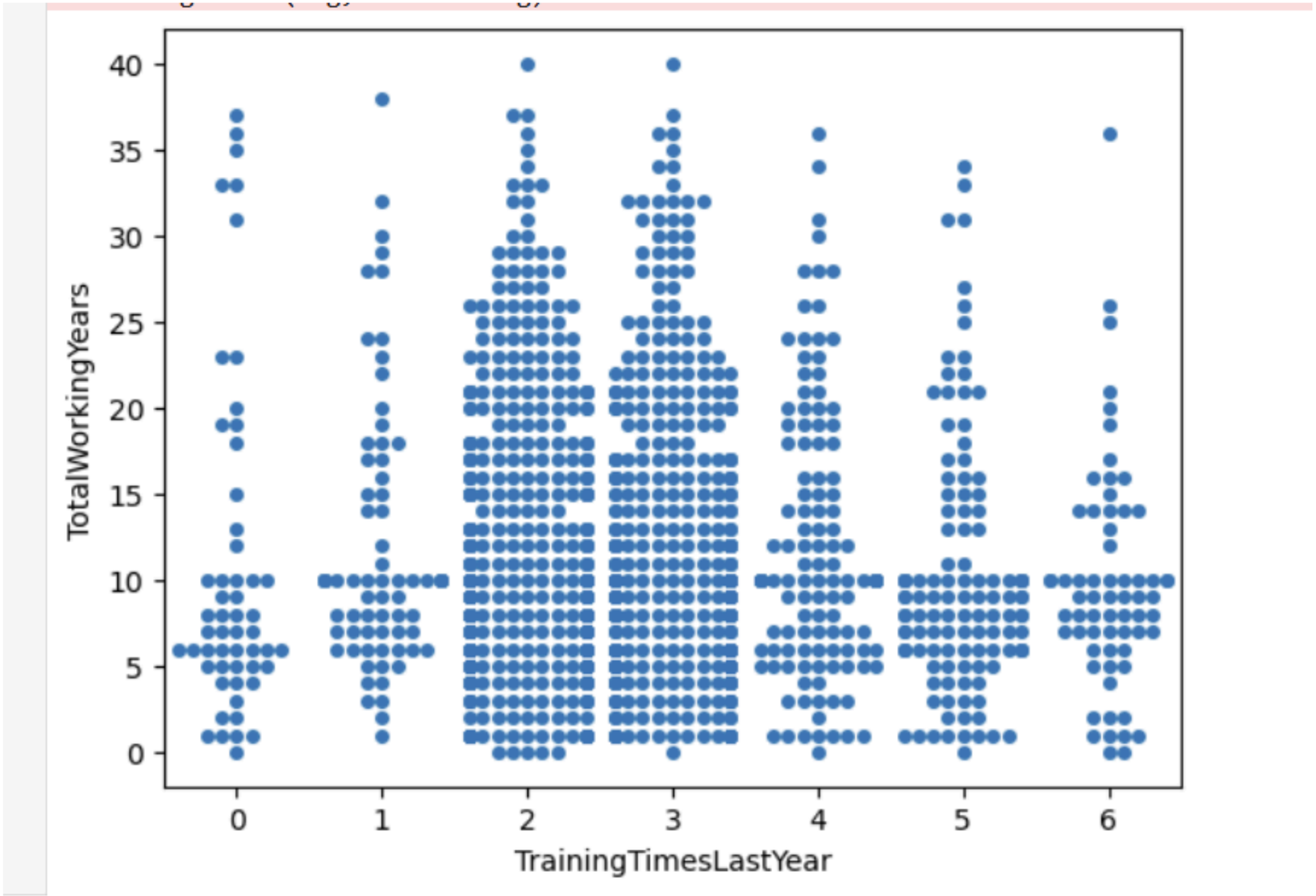

```
[ ]:
```

Impact Of Environment Satisfaction On Total Working Years

```
warnings.warn(msg, UserWarning)
```



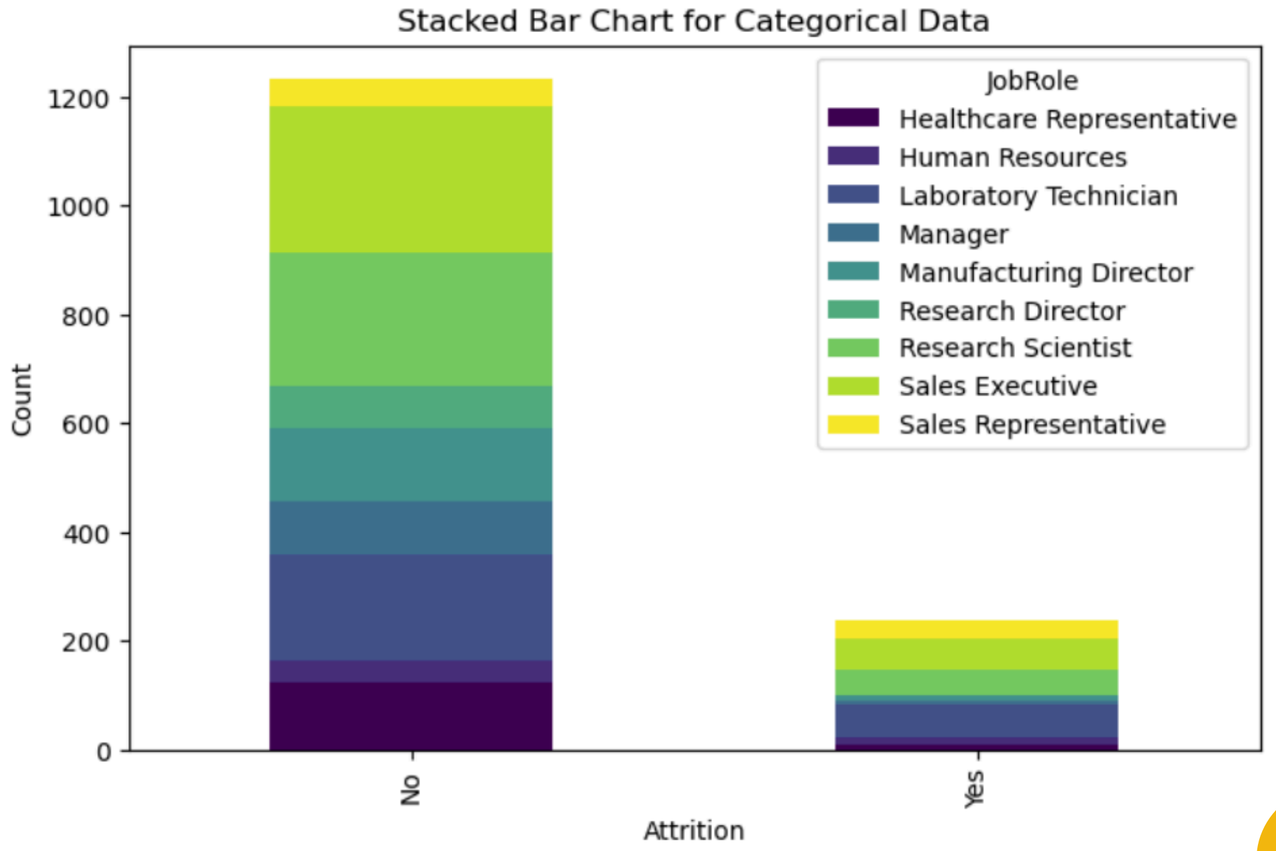
Impact Of Training Time On Total Working Years



Attrition In Each Job role

[ ]:
```
# Attrition Vs JobRole
```

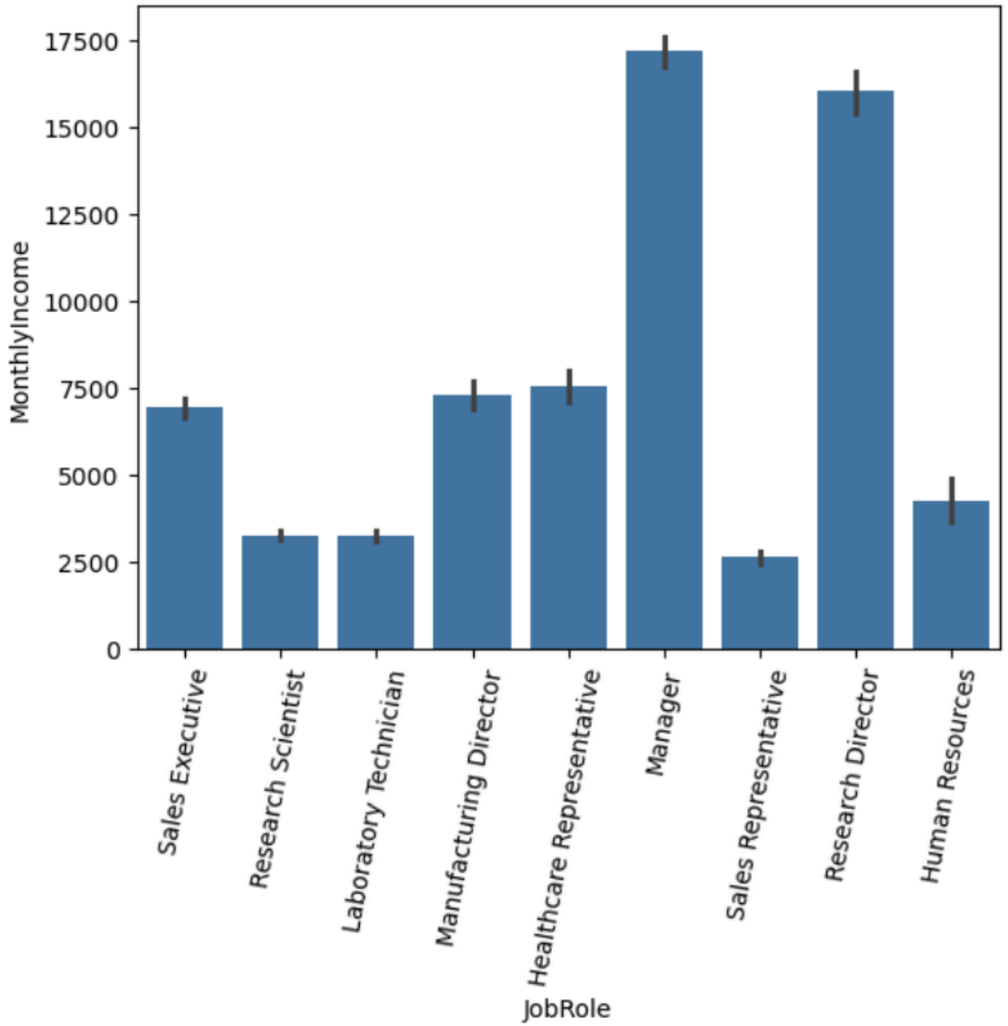
[70]:
```
df_grouped = hr.groupby(['Attrition', 'JobRole']).size().unstack()

df_grouped.plot(kind='bar', stacked=True, colormap='viridis', figsize=(8,5))
plt.ylabel("Count")
plt.title("Stacked Bar Chart for Categorical Data")
plt.show()
```



Average Monthly Income In Each Job Role

```
[ ]:
```

```
[72]: sns.barplot(data=hr,x='JobRole',y='MonthlyIncome')
plt.xticks(rotation=80)
plt.show()
```



Conclusion

The analysis of employee attrition provided valuable insights into the factors influencing workforce turnover. The key findings reveal trends related to gender, department, experience, salary, and job roles, helping organizations understand employee retention challenges.

🔹 Key Insights & Possible Reasons:

1️⃣ Gender Disparity in Attrition: While there are more male employees in the organization, their attrition rate is also higher, indicating potential concerns related to job satisfaction or workplace policies affecting male employees more significantly.

2️⃣ Overall Attrition Rate: The company's overall attrition rate is 16%, suggesting a moderate level of employee turnover that may require strategic retention initiatives.

3️⃣ Department-wise Attrition: The R&D department has a higher attrition rate compared to others. One possible reason is that frequent business travel is more common in this department, which might be impacting job satisfaction and work-life balance.

4️⃣ Early Career Attrition: Employees in their first 1 to 4 years at the company showed higher attrition rates, possibly due to lower salaries and limited career growth opportunities in the early stages.

5 Impact of Salary on Attrition: The average monthly income of employees who left was significantly lower than those who stayed, reinforcing the idea that salary dissatisfaction is a major driver of attrition.

6 Effect of Business Travel: While frequent travel was observed in certain departments, overall travel frequency had no significant impact on company-wide attrition, indicating that other factors play a bigger role in employee turnover.

7 Experience & Job Satisfaction: Employees with less experience exhibited higher attrition rates, possibly due to lower job satisfaction and dissatisfaction with the work environment, which could lead to early exits.

8 High Attrition in Sales Representatives: Among different job roles, Sales Representatives had the highest attrition rate. One possible reason is their low salary, which may be discouraging long-term retention.

Final Thoughts:

This analysis highlights that salary dissatisfaction, early-career retention challenges, work-life balance, and job role expectations are some of the main factors influencing attrition. Organizations should focus on competitive compensation, employee engagement, career development programs, and better job satisfaction initiatives to reduce turnover and retain talent effectively.

Final Thoughts:

This analysis highlights that salary dissatisfaction, early-career retention challenges, work-life balance, and job role expectations are some of the main factors influencing attrition. Organizations should focus on competitive compensation, employee engagement, career development programs, and better job satisfaction initiatives to reduce turnover and retain talent effectively.

Thank you

dimpybangoriya@gmail.com